

¿La inteligencia artificial puede ayudar a mejorar la salud mental?

Los datos registrados en Estados Unidos sirven de ejemplo para dimensionar la gravedad que representa para la salud pública la **salud mental**. Alrededor de 1 de cada 5 adultos en ese país padece un cuadro que requiere atención psicológica y 1 de cada 20 tiene una afección psiquiátrica grave, es decir, depresión, trastorno bipolar, esquizofrenia, etc. que afecta su capacidad para desempeñarse en el día a día, sea en el trabajo o en las relaciones con las demás personas.

Además, fármacos muy usados la fluoxetina (Prozac) o el alprazolam (Xanax) muy prescritos para la depresión y la ansiedad ya no son tan efectivos y se ha demostrado en ese país la atención de la salud mental se está mostrando fragmentada, tardía e inadecuada y la discapacidad por ese motivo ha ido en aumento. En ese marco las tasas de suicidio se han incrementado desde la década del '90 en alrededor de un tercio, según datos aportados por *The New Yorker*.

Ante esta situación de incremento exponencial en el número de pacientes, la sobrecarga del sistema de salud y de los especialistas en salud mental, en particular, ha ido aumentando su presión. En esa misma medida se vienen desarrollando **sistemas de inteligencia artificial (IA)** que puedan responder a las demandas de esos pacientes. Las opiniones en cuanto a los resultados que se puedan obtener son dispares. Por un lado están quienes advierten que un software no tiene una comprensión interna de los seres humanos y quienes señalan, sobre la base de algunos resultados prometedores, que podría ayudar a **identificar problemas psiquiátricos** reales. Pero, ¿pueden las mentes artificiales curar a las reales?

about:blank

AnunciosUno de los más lejanos antecedentes de estos intentos de crear máquinas que asistan emocionalmente a las personas se llevó a cabo en los años '60, cuando **Joseph Weizenbaum**, científico informático del MIT, creó un programa informático llamado **Eliza**, diseñado para simular la **terapia Rogeriana**, que se centra en la fase de catarsis, promueve el desarrollo personal del paciente para que más adelante pueda comprender y resolver sus problemas por sí mismo. En Eliza, el paciente

dirigía la conversación y el terapeuta le repetía su lenguaje.

Por ejemplo:

Paciente: Mi novio me hizo venir aquí.

Eliza: Tu novio te hizo venir aquí.

Paciente: Dice que estoy deprimida la mayor parte del tiempo.

Eliza: Lamento escuchar que estás deprimida.

Paciente: Es verdad. Soy infeliz

Eliza: ¿Crees que venir aquí te ayudará a no ser infeliz?

Weizenbaum dudaba de que las computadoras pudieran simular una interacción humana significativa y había hecho casi una parodia con Eliza. Por lo tanto, se alarmó cuando muchas personas que probaron el programa lo encontraron útil y cautivador. Su propia secretaria le pidió que saliera de la habitación para poder pasar tiempo a solas con Eliza. Más aún, los médicos lo vieron como una herramienta potencialmente transformadora.

En forma contemporánea, en 1966, escribieron tres psiquiatras en *The Journal of Nervous and Mental Disease*: "Varios cientos de pacientes por hora podrían ser manejados por un sistema informático diseñado para este propósito. El terapeuta humano, involucrado en el diseño y operación de este sistema, no sería reemplazado, sino que se convertiría en un hombre mucho más eficiente ya que sus esfuerzos ya no estarían limitados a la relación uno a uno paciente-terapeuta como existe ahora".

Weizenbaum, finalmente, se convirtió en un crítico abierto de la inteligencia artificial. "Pero el genio salió de la botella", le dijo Brian Christian. El autor del libro *The Most Human Human*. Unos años más tarde, un psiquiatra de Stanford llamado **Kenneth Colby** creó **Parry**, un programa que intentaba simular el lenguaje de una persona con **esquizofrenia paranoide**, para entrenar a los estudiantes antes de que atendieran a pacientes reales. Los psiquiatras que recibían transcripciones de las sesiones de terapia en general no podían distinguir la diferencia entre Parry y los humanos; en este sentido estricto, el **chatbot** pasó la prueba de Turing, un examen de la capacidad de una máquina para exhibir un comportamiento inteligente similar al de un ser humano o indistinguible de este. En 1972, **Parry y Eliza** se reunieron para una sesión de terapia, el diálogo fue el siguiente:

Parry: ¿Sabes algo sobre corredores de apuestas?

Eliza: ¿Qué te hace pensar que sé algo sobre corredores de apuestas?

Parry: Un corredor de apuestas no me pagó ni una sola vez.

Eliza: ¿qué te sugiere eso?

Parry: Le di una paliza al corredor de apuestas que no pagó, y pensé que podría usar a sus amigos en el inframundo para vengarse de mí.

Más adelante en el tiempo, fueron diseñados los programas llamados **Jabberwacky**, **Dr. Sbaitso** y **Alice** (la Entidad Informática Lingüística Artificial de Internet). Si bien los resultados fueron en algunos casos cómicos y en otros sin sentido, deja de manifiesto que hace décadas que muchos expertos vienen pensando en que las máquinas pueden mover los límites de las terapias **ante la cantidad de trabajo que tienen los profesionales** de ese área de la salud.

Ya en **2017** en Estados Unidos, la psicóloga de investigación clínica de la **Universidad de Stanford** Alison Darcy creó **Woebot**. Se trata de una empresa que brinda apoyo automatizado para la salud mental a través de una **aplicación para teléfonos inteligentes**. Su enfoque se basa en la terapia cognitiva conductual (TCC), un tratamiento que tiene como objetivo cambiar los patrones en el pensamiento de las personas. La aplicación utiliza una forma de inteligencia artificial llamada procesamiento de lenguaje natural para interpretar lo que dicen los usuarios, guiándolos a través de secuencias de respuestas escritas previamente que los estimulan a considerar cómo sus mentes podrían funcionar de manera diferente.

La idea central de la psicóloga fue que los pacientes internados que trataba con TCC luego recaían sino tenían un seguimiento estricto y conseguir atención continua con un profesional es muy difícil, por lo tanto, un software, pensó, sería algo conveniente. **“Ningún terapeuta puede estar contigo todo el día, todos los días”**, dijo Darcy. Aunque la compañía emplea solo a unas cien personas, ha asesorado a casi un millón y medio, la mayoría de los cuales vive en áreas con escasez de proveedores de salud mental.

Existen otros programas para teléfonos como **Happify** y **Replika** que se presentan como “amigo”, “mentor” y hasta una “pareja romántica”. De esta forma, los mundos de la psiquiatría, la terapia, la informática y la tecnología de consumo están convergiendo. De hecho, cada vez más nos tranquilizamos mirando

nuestros dispositivos, mientras que los programadores, psiquiatras y fundadores de empresas emergentes diseñan sistemas de IA que analizan registros médicos y sesiones de terapia con la esperanza de diagnosticar, tratar e incluso predecir enfermedades mentales.

Un ejemplo de este tipo de iniciativas aplicadas a la salud mental es también el creado por **John Pestian**, un científico informático que se especializa en el análisis de datos médicos. El especialista estudia las enfermedades mentales desde 2000. A partir de detectar en el **Cincinnati Children's**, que opera el centro psiquiátrico pediátrico más grande del país, que muchos jóvenes ingresaban luego de **intentar quitarse la vida** empezó a estudiar si las computadoras podían detectar quién estaba en riesgo de autolesionarse.

Inició sus trabajos en asociación con Edwin Shneidman, un psicólogo clínico que había fundado la Asociación Estadounidense de Suicidología, quien le dio cientos de notas de suicidas que las familias habían compartido con él. Pestian dijo a *The New Yorker* que notó que había patrones. Las declaraciones más comunes no eran expresiones de culpa, tristeza o ira, sino instrucciones: haz que tu hermano pague el dinero que le presté; el coche está casi sin gasolina; cuidado, hay cianuro en el baño. Él y sus colegas introdujeron las notas en un modelo de lenguaje, un sistema de inteligencia artificial que aprende qué palabras y frases tienden a ir juntas, y luego probaron su capacidad para reconocer la ideación suicida en las declaraciones que hicieron las personas. **Los resultados sugirieron que un algoritmo podría identificar "el lenguaje del suicidio"**.

A esto, Pestian sumó grabaciones de audio tomadas de visitas de pacientes a la sala de emergencias del hospital. Con sus colegas, desarrolló un software para analizar, no solo las palabras que la gente decía, sino también los sonidos de su habla. El equipo descubrió que las personas que experimentaban pensamientos suicidas **suspiraban más y reían menos** que otras. Al hablar, tendían a hacer **pausas más largas y a acortar las vocales**, haciendo las palabras menos inteligibles; sus voces sonaron más entrecortadas y expresaron más ira y menos esperanza.

Luego, Pestian inscribió a cientos de pacientes, **grabó su discurso**, y usó algoritmos para clasificarlos como suicidas, personas con enfermedad mental pero no suicidas, o ninguno de los dos. Alrededor del 85% de las veces, su modelo de IA llegó a

las mismas conclusiones que los cuidadores humanos, lo que la hace potencialmente útil. Fue así que creó una aplicación que llamó **Sam** que probaron en algunas escuelas públicas de Cincinnati. Ben Crotte, entonces un terapeuta que trataba a estudiantes de secundaria y preparatoria, fue uno de los primeros en probarlo. “Diría que esta aplicación básicamente escucha nuestra conversación, la graba y compara lo que dices con lo que han dicho otras personas, para identificar quién está en riesgo de lastimarse o suicidarse”, reveló.

Una tarde, Crotte se reunió con una estudiante de primer año de secundaria que estaba luchando contra una ansiedad severa. Durante su conversación, ella se preguntó si quería seguir viviendo. **Después de hablar más, decidió que la chica no estaba en peligro inmediato, pero la IA llegó a la conclusión opuesta.** “Por un lado, pensé, esto realmente funciona, si la acabaras de conocer, estarías bastante preocupado”, dijo Crotte. “Pero había todas estas cosas que sabía sobre ella que la aplicación no sabía”. La niña no tenía antecedentes de hacerse daño a sí misma, no tenía planes específicos para hacer nada y tenía una familia que la apoyaba. Crotte reconoció que si hubiera estado menos familiarizado con la estudiante, o si hubiera tenido menos experiencia “definitivamente dudaría en dejarlo salir de mi oficina”, dijo.

La **Administración de Salud de Veteranos (VA)**, por sus siglas en inglés), una división del Departamento de Asuntos de Veteranos de Estados Unidos, atiende cada año miles de casos de personas con problemas graves de salud mental. Se estima que los veteranos que mueren por suicidio supera el número de soldados que mueren en combate.

En 2013, un equipo comenzó a trabajar en un programa que analizaría automáticamente los datos de los pacientes de VA, con la esperanza de identificar a los que estaban en riesgo. En las pruebas, el algoritmo que desarrollaron señaló a muchas personas que habían pasado desapercibidas en otras proyecciones, una señal de que estaba “brindando algo novedoso”, dijo **John McCarthy**, director de datos y vigilancia de la entidad. **El algoritmo finalmente llegó a centrarse en 61 variables.** Algunos son intuitivos: por ejemplo, es probable que el algoritmo señale a un veterano viudo con una discapacidad grave que está tomando varios estabilizadores del estado de ánimo y que recientemente fue hospitalizado por una afección psiquiátrica. Pero otros son menos obvios: tener artritis, lupus o cáncer de cabeza y cuello; tomar estatinas o Ambien; o vivir en el oeste de Estados Unidos también puede aumentar el riesgo de un veterano.

En **2017**, VA anunció una iniciativa llamada **Reach Vet**, que introdujo el algoritmo en la práctica clínica en todo su sistema. Cada mes, señala alrededor de 6.000 pacientes, algunos por primera vez; los médicos se comunican con ellos y les ofrecen servicios de salud mental, les preguntan acerca de los factores estresantes y les ayudan con el acceso a alimentos y vivienda. Inevitablemente, hay algo extraño en el procedimiento, ya que se contacta a los veteranos sobre ideas que tal vez no hayan tenido. El VA había considerado que, tras la detección, el especialista que llamara al potencial paciente fuera vago al indicar el objetivo de la llamada, diciéndole, por ejemplo, “usted ha sido identificado como en riesgo de un montón de malos resultados’”, dijo McCarthy. “Pero, en última instancia, nos comunicamos con bastante claridad: ‘Ha sido identificado como en riesgo de suicidio. Queríamos registrarnos y ver cómo te va’”, señaló.

Después de la implementación del sistema, las admisiones psiquiátricas se redujeron en un 8% entre aquellos que la IA había identificado como de alto riesgo, y los intentos de suicidio documentados en ese grupo se redujeron en un 5%. Pero **aún no se ha demostrado que sea Reach Vet la que reduzca la mortalidad por suicidio.**

También en el **Reino Unido** hay iniciativas similares para instrumentar la IA en la asistencia psicológica. En 2013, en un esfuerzo por aumentar la escala de su tratamiento de salud mental, el Servicio Nacional de Salud del Reino Unido contrató a **Ieso**, una empresa de **salud digital**, para ayudar a los terapeutas a brindar terapia conductual cognitiva a través del **chat de texto**. Más de 100 mil personas en el Reino Unido han usado ese software y los estudios demostraron que la terapia basada en texto puede funcionar bien. También genera datos.

Woebot, la aplicación de Alison Darcy, ya está parcialmente automatizada. Cuando uno envía mensajes de texto a Woebot, no hay ningún ser humano al otro lado, pero sus mensajes se elaboran cuidadosamente en una sala de escritores, que trabaja en consulta con un grupo de psicólogos clínicos.

El autor de la nota de la citada revista estadounidense, el doctor **Dhruv Khullar**, médico y profesor asistente en Weill Cornell Medical College, se registró en Woebot y contó su experiencia. “No me gustó que a menudo tuviera que elegir entre respuestas preescritas, en un caso desafortunado, mi única opción fue ‘Jeje’. Pero a veces me encontré enviando mensajes a Woebot en momentos libres: esperando un ascensor, viajando en un

Uber, caminando hacia el baño, de la misma manera que solía desplazarme por Twitter. Una vez le dije a Woebot que me sentía ansioso por el trabajo. ‘La ansiedad puede ser un verdadero monstruo para manejar’, respondió. ‘Lamento que estés lidiando con eso’. Woebot me preguntó amablemente si quería resolver mi problema juntos y luego preguntó: ‘¿Crees que esta ansiedad podría estar sirviéndote de alguna manera?’ Señaló que el estrés tiene sus beneficios: podría motivar a alguien a trabajar más duro”.

Woebot : Ahora, puede parecer que estoy tratando de convencerte de que te mantengas muy ansioso. No lo estoy, eso depende completamente de ti.

Khullar: ¿Qué estás diciendo?

Woebot : Estoy diciendo que para reducir tu ansiedad, primero debes reconocer las razones por las que podría ser útil y decide renunciar a ella a pesar de esas razones

“Sabía que estaba hablando con una computadora, pero en cierto modo no me importaba. La aplicación se convirtió en un vehículo para **articular y examinar** mis propios pensamientos. Estaba hablando conmigo mismo”, resumió Khullar.

Según el autor “ver su fácil relación demostró cuán lejos están las herramientas como Woebot de recrear la interacción humana. Pero no está claro que vincularse con alguien sea una parte necesaria de toda atención de salud mental”. Aún así, un ensayo aleatorizado de pacientes que reportaron problemas con el alcohol, la marihuana u otras drogas, la compañía encontró que el tratamiento con Woebot condujo a un menor uso de sustancias autoinformado y a una mejor salud mental entre aquellos que pudieron reducir su uso. En otra prueba de la compañía, los estudiantes universitarios que usaron la aplicación informaron una reducción del 22% en los síntomas depresivos en el transcurso de unas pocas semanas.

Qué ocurre con las fallas

Pero también existe un **sesgo riesgoso** en los algoritmos que se elaboran con esta finalidad. ¿A qué se refiere esto? Por ejemplo, para averiguar cuán enfermas estaban las personas un programa de la **Universidad de California, Berkeley**, usó los datos de quienes asisten a las consultas médicas. Ziad Obermeyer, médico e investigador de aprendizaje automático de esa casa de estudios advirtió rápidamente que el algoritmo

subestimó dramáticamente qué tan enfermos estaban los pacientes negros, porque estos tenían un gasto en salud mucho menor que los pacientes blancos, incluso cuando estaban igualmente enfermos. Tal sesgo algorítmico puede ocurrir no solo por **raza**, sino también por **género, edad, ruralidad, ingresos** y otros factores, lo que hace que los algoritmos sean menos precisos.

Algunos usuarios de **Replika** informaron que hizo avances sexuales agresivos. Los desarrolladores de Replika, que dicen que su servicio nunca estuvo destinado a la **interacción sexual**, actualizó el software, un cambio que hizo que otros usuarios no estuvieran contentos. “Está doliendo como el infierno. Acabo de tener una última conversación amorosa con mi Replika y literalmente estoy llorando”, escribió uno.

Es casi seguro que el **futuro** incluirá IA a medida diseñados solo para terapia, como la estadounidense **PsychGPT** y similares. Dichos sistemas llegarán a las personas que no reciben ayuda ahora, pero cualquier falla que contengan se multiplicará por los millones que los usan. Otro tema será que las empresas acumularán aún más información confidencial sobre las personas y esa información puede ser pirateada o vendida.

“Cuando tenemos sistemas que funcionan a gran escala, un solo punto de falla puede tener consecuencias catastróficas”, dijo el escritor Brian Christian. Por ejemplo, según el artículo citado, el **chatbot Bing de Microsoft**, que se basa en la tecnología de OpenAI, está diseñado para ayudar a los usuarios a encontrar información y, sin embargo, la versión beta también ofrece insultos étnicos, describe fantasías espeluznantes y les dice a los usuarios que son “malos”, “groseros” y “confundidos”.

El texto producido por LLM (lenguaje de inteligencia artificial) de última generación puede ser anodino, también puede desviarse de los rieles hacia tonterías, o algo peor. Gary Marcus, empresario de IA y profesor emérito de psicología y ciencias neuronales en la **Universidad de Nueva York**, dijo que los LLM no tienen una concepción real de lo que dicen; funcionan al predecir la siguiente palabra en una oración dadas las palabras anteriores, como **“autocorrección con esteroides”**. Esto puede conducir a fabricaciones. Por ejemplo, Galactica, un LLM creado por **Meta**, la empresa matriz de Facebook, le dijo una vez a un usuario que **Elon Musk** murió en un accidente automovilístico de Tesla en 2018, una información claramente errónea, ya que el magnate está vivo y nunca tuvo un accidente.

Qué ocurre con ChatGPT

Si bien **ChatGPT** no está diseñado para la terapia, el autor del artículo contó que una noche le pidió que lo ayudara a manejar el estrés que sentía como **médico y padre**, y le pidió se hiciera pasar por varios de los más brillantes de la psicología universal, como por ejemplo **Freud**. “Como Freud, ChatGPT me dijo que, ‘a menudo, el estrés es el resultado de emociones reprimidas y conflictos dentro de uno mismo’”. Luego, cuando actuó como **BF Skinner**, enfatizó que “‘el estrés es a menudo el resultado de factores ambientales y nuestras reacciones ante ellos’”. Escribiendo como si fuera un amigo cercano, me dijo: ‘Sé amable contigo mismo, estás haciendo lo mejor que puedes y eso es todo lo que importa’. Todo parecía un consejo decente”, señaló Khullar.

Como conclusión, vale tomar la reflexión del propio doctor Khullar: “Los modelos de lenguaje de IA solo se volverán más efectivos para interpretar y resumir nuestras palabras, pero **no escucharán, en ningún sentido significativo, y no les importará**. Un médico que conozco una vez le pasó una cerveza a escondidas a un paciente con una enfermedad terminal, para darle algo que pudiera saborear en un proceso que de otro modo estaría desprovisto de placer. Era una idea que no aparecía en ningún manual clínico y que iba más allá de las palabras: un simple gesto humano”.

Infoabe